УДК 004.933

Ю.В. Крак[1], О.В. Бармак[2], В.С. Касьянюк[1], І.О. Стеля[1]
[1]Київський національний університет імені Тараса Шевченка, Україна
вул. Володимирська, 60, м. Київ, 01601
[2]Хмельницький національний університет, Україна
вул. Інститутська, 11, м. Хмельницький, 29016

МОДЕЛІ ПОДАННЯ ГРАМАТИЧНИХ КОНСТРУКЦІЙ ДЛЯ АВТОМАТИЗОВАНОГО ПЕРЕКЛАДУ ТЕКСТОВОЇ ІНФОРМАЦІЇ УКРАЇНСЬКОЮ МОВОЮ НА ЖЕСТОВУ МОВУ

Iu.V. Krak[1], O.V. Barmak[2], V.S. Kasianiuk[1], I.O. Stelia[1]
[1]Taras Shevchenko National University of Kyiv, Ukraine
60, Volodymyrska str., Kyiv, 01601
[2]Khmelnytskyi National University, Ukraine
11, Instytutska str., Khmelnytskyi, 29026

MODELS OF GRAMMAR CONSTRUCTIONS REPRESENTATION FOR AUTOMATED TRANSLATION TEXTS IN UKRAINIAN INTO GESTURES LANGUAGE

На основі теоретико-множинної моделі української мови запропоновано метод автоматизованого перекладу текстової інформації з української мови на жестову мову шляхом отримання взаємно однозначних пар конструкцій, що передають значення речення на вхідній мові у відповідне речення на жестовій мові. Побудовано моделі подання таких граматичних конструкцій, розроблено і алгоритмічно та програмно реалізовано відповідну інформаційну технологію автоматизованого перекладу.

Ключові слова: теоретико-множинна модель, автоматизований переклад, жестова мова, граматична конструкція

Method of text information automated translation from Ukrainian into gesture language on basis of set-theoretical model of Ukrainian by receipting mutually bijective constructions pairs that pass sense has been suggested: sentence in input language into corresponding sentence in gesture language. The models of presentation of such grammatical constructions are built corresponding information technology of automated translation is worked out, algorithmically and programmatically realized.

Key words: set-theoretical model, automated translation, gesture language, grammatical construction

Вступ

Співвідношення кількості нечуючих людей і людей з важкими формами втрати слуху до кількості людей, які чують, у всьому світі в середньому складає приблизно один до тисячі – це десятки мільйонів людей, для яких необхідно створювати засоби рівноцінного спілкування у суспільстві. Важливою на сьогодні є проблема створення сучасних інформаційних технологій, що включають алгоритми, за допомогою яких можна було б створювати нові комп’ютерні системи навчання та комунікації для людей з вадами слуху. Для вирішення цієї проблеми запропонована концепція інформаційної технології невербального спілкування людей з вадами слуху [1,2]. Комплексна інформаційна технологія включає в себе функціональність за синтезом: рухів жестової мови, дактильної абетки, міміки промовляння на тривимірній моделі людини та аналізу невербальних каналів комунікації.

Одним із напрямків концепції є створення автоматизованої системи перекладу з української на жестову мову. Жестова мова – це самостійна мова, яка виникла природно та складається з комбінації жестів, кожен з яких виконується руками в

поєднанні з мімікою, формою або рухом рота і губ, а також у поєднанні з положенням корпусу тіла. Жести, умовно-схематичні, іноді придумуються на льоту, необов'язково мають візуальний зв'язок з позначуваним словом. Також вони не є візуальною інтерпретацією звичайних мов; вони мають свою граматику, можуть бути використані для обговорення самих різних тем: від простих і конкретних до піднесених або абстрактних [3].

Лексика розмовної жестової мови ще недостатньо вивчена, особливо ідіоматика, фразеологія, морфологія. Автоматичне приписування розмовній жестовій мові всіх форм словесної та письмової мови є неправильним. Не можна втиснути розмовну жестову мову глухих у традиційні моделі лінгвістики. Для опису фактів морфології розмовної жестової мови найбільше підходить опис за принципом: від значення – до форми. Значення «множинності», «часу», «приналежності», «закінчення дії» передаються у жестовому мовленні своїми способами. Залежно від значення змінюється і спосіб виконання жесту (якість руху).

Система автоматизованого перекладу з української мови на жестову мову передбачає створення множинної моделі словника української мови та реалізації її для можливості моделювання відповідних пов’язаних узагальнених граматичних конструкцій української та жестової мов. До розгляду беруться тільки прості речення. Вважатимемо, що довільне складне речення можна подати як декомпозицію простих речень. Підмет і присудок у реченнях пов’язані предикативним зв’язком. Під простими реченнями розумітимемо речення з одним предикативним зв’язком. Як для української, так і для жестової мови обмежимося такими типами простих речень: розповідними (стверджувальними, окличними та заперечними), питальними та спонукальними [4].

Для реалізації інформаційної технології автоматизованого перекладу запропоновані відповідні інфологічні моделі: для словників української та жестової мов; для подання пов’язаних узагальнених граматичних конструкцій для перекладу.

Постановка задачі

Метою даної роботи є розробка інформаційної технології перекладу з української на жестову мову. Для реалізації цієї мети потрібно вирішити наступні завдання: 1) створити інфологічне подання словників української та жестової мов; 2) розробити механізм подання відношень між словами української та жестами жестової мов; 3) розробити модель подання узагальнених граматичних конструкцій для української та жестової мов; 4) розробити інформаційну технологію для здійснення автоматизованого перекладу за допомогою отриманих моделей; 5) виконати алгоритмічну та програмну реалізацію інформаційної технології для підтвердження її життєздатності.

Інфологічні моделі словників української та жестової мов

Для досягнення мети перекладу необхідно запропонувати граматичний словник української мови. Українська мова відноситься до флективних мов – мов синтетичного типу, у якому домінує словозміна за допомогою флексій – формантів, що поєднують відразу декілька значень. Флективна будова мови протистоїть аглютинативній, у якій кожен формант несе тільки одне значення. Для розробки граматичного словника візьмемо теорію лексикографічних систем [5].

Побудова граматичного словника флективної мови визначається наявністю формальної моделі словозміни, що означає встановлення та формалізацію лінгвістичних критеріїв, згідно з якими вся множина слів мови розбивається на певні підмножини, взаємний перетин яких є порожнім, а всередині кожної з них

словозміна відбувається за однаковими правилами. Підмножини слів з такими властивостями називаються словозмінними парадигматичними типами.

Під парадигматичним типом розумітимемо групу лексем, словозмінна парадигма яких характеризується однаковою кількістю граматичних форм, усередині якої словозміна відбувається за тим самим (єдиним) правилом. Для української мови, яка є мовою аналітико-синтетичного типу, це означає, що, по-перше, слова, які належать до одного парадигматичного класу, мають однакові флексії у відповідних граматичних значеннях та однаковий характер чергування в основі і, по-друге, відповідні аналітичні форми будуються за однаковими моделями їх утворення).

Українській мові притаманні наступні граматичні категорії, що визначають словозміну: s – число (однина(s_1), множина(s_2)), g – рід (чоловічий (g_1), жіночий (g_2), середній (g_3), вищий ступінь (g_4), чоловічий або середній рід для істоти (g_5), жіночий або середній рід для істоти (g_6),чоловічий або жіночий рід для істоти (g_7), незмінювана словникова одиниця (g_8), найвищий ступінь (g_9), i – узагальнена категорія, що включає відмінок, особу, час, стан, спосіб (називний (i_1), родовий (i_2), давальний (i_3), знахідний (i_4), орудний (i_5), місцевий (i_6), кличний (i_7), інфінітив (i_8), 1 особа, наказовий спосіб (i_9), 2 особа, наказовий спосіб (i_{10}), 3 особа, наказовий спосіб (i_{11}), 1 особа, майбутній час (i_{12}), 2 особа, майбутній час (i_{13}), 3 особа, майбутній час (i_{14}), 1 особа, теперішній (майбутній) час (i_{15}), 2 особа, теперішній (майбутній) час (i_{16}), 3 особа, теперішній (майбутній) час (i_{17}), 1 особа, теперішній час (i_{18}), 2 особа, теперішній час (i_{19}), 3 особа, теперішній час (i_{20}), активний дієприкметник, теперішній час (i_{21}), дієприслівник, теперішній час (i_{22}), минулий час (i_{23}), активний дієприкметник, минулий час (i_{24}), пасивний дієприкметник, минулий час (i_{25}), безособова форма, минулий час (i_{26}), дієприслівник, минулий час (i_{27}), незмінювана одиниця (i_{28})).

Кожне слово української мови віднесемо до наступних класів (частин мови) (p): іменники (p_1), числівники порядкові (p_2), числівники кількісні (p_3), числівники (p_4), числівники типу «два» (p_5), дієслова недоконаного і доконаного виду (p_6), дієслова доконаного виду (p_7), дієслова недоконаного виду (p_8), дієприкметники (p_9), прикметники (p_{10}), займенники (p_{11}), займенники-прикметники (p_{12}), прислівники (p_{13}), вигуки (p_{14}), сполучники (p_{15}), частки (p_{16}), прийменники (p_{17}), присудкові слова (p_{18}), вставні слова (p_{19}), абревіатури (p_{20}), сполуки (p_{21}), дієприслівники (p_{22}), займенники з часткою (p_{23}), сполучники і частки (p_{24}), прислівники і частки (p_{25}), числівники з прийменником (p_{26}), іменники з прийменником (p_{27}), дієприслівники з часткою (p_{28}), займенники з прийменником (p_{29}), дієприслівники з часткою (p_{30}).

Для української мови введемо наступні парадигматичні типи (табл. 1).

Виходячи з того, що українська мова є флективною мовою (тобто граматичні значення передаються флексіями), слова мови моделюються у вигляді комбінації незмінної та змінної складових:

$$x = c(x)\,\&\,f(x), \qquad (1)$$

де $c(x)$ – частина лексеми x, яка в процесі словозміни залишається незмінною (квазіоснова), $f(x)$ – її змінна складова (квазіфлексія), & – конкатенація.

Таблиця 1. Парадигматичні типи української мови

Парадигма-тичні типи	Граматичні класи	Граматичні категорії, що визначають словозміну	Кількість граматичних значень
T^1	p_1	$s*(i_1,\ldots,i_7)$	14
T^2	$p_2 \cup p_4 \cup p_9 \cup p_{10} \cup p_{12}$	$(g_1,g_2,g_3)*((s_1,s_2)*(i_1,\ldots,i_7))$	24
T^3	$p_3 \cup p_{11}$,	$(i_1,\ldots,i_6)$	6
T^4	p_5	$(g_1,g_2)*(i_1,\ldots,i_6)$	12
T^5	$p_6 \cup p_8$	$i_8,(i_{10}*s_1),(i_9*s_2),(i_{10}*s_2),(i_{12}*s_1),$ $((i_{12},\ldots,i_{17})*s),i_{21},i_{22},$ $i_{23}*(g_1,g_2,g_3)*s_1),$ $(i_{23}*s_2),i_{24},i_{25},i_{26},i_{27}$	26
T^6	p_7	$i_8,(i_{10}*s_1),(i_9,s_2),(i_{10},s_2),$ $((i_{12},i_{13},i_{14})*s),$ $(i_{23}*(g_1,g_2,g_3)*s_1),$ $(i_{23}*s_2),i_{24},i_{25},i_{26},i_{27}$	18
T^0	$\bigcup_{j=13}^{30} p_j$	i_{28}	1

Реалізація такої моделі (1) зберігання слова прибирає надмірну надлишковість, маючи на увазі, що в українській мові декілька мільйонів слів (у всіх словозмінах) та тільки біля 100 тисяч базових слів – інфінітивів. Тобто, замість повного тексту слова у таблиці зберігається номер слова-інфінітива із відповідної множини інфінітивів, номер позиції у слові, до якої слово-інфінітив незмінне, та номер флексії (змінної частини слова) із відповідної множини. Тобто, слова української мови подаються у вигляді:

$$W = \{W_i : W_i = \{I_{i_1} \in I, F_{i_2} \in F, k, g \in G, s \in S, in \in In\}\}, \qquad (2)$$

де W_i – параметри слова української мови ($i = 0,\ldots,N-1$, N – кількість слів у словнику); F – множина всіх можливих закінчень слів (флексій); k – номер позиції у слові-інфінітиві, з якої починається конкатенуватися флексія (можливі випадки коли $k = 0$ – для словоформи, яка повністю відміна від інфінітива); $G = \{g_1,\ldots,g_9\}$; $S = \{s_1, s_2\}$; $In = \{i_1,\ldots,i_{28}\}$, I – множина слів-інфінітивів української мови (для дієслів – інфінітиви, для іменників – слова у називному відмінку, однині тощо):

$$I = \{I_i : I_i = \{word\inf, p \in P\}\},\qquad (3)$$

де $P \in \{p_1, \ldots, p_{30}\}$; $word\inf$ – слово-інфінітив.

Формування множин W, F, I проходить наступним чином (для всіх слів-інфінітивів мови):

- *Вхід*: слово-інфінітив та усі його словоформи (граматичні значення)
- $k =$ Довжина(загальна частина для слова-інфінітива та всіх його словоформ)
- Для всіх словоформ визначаємо закінчення (відмінна частина) та додаємо його до множини F (якщо такого закінчення там немає)
- Додаємо слово-інфінітив до множини I (3)
- Формуємо множину W (2)

Прямою задачею для такої моделі (2) є проста задача відтворення повного тексту слова за трьома параметрами I_{i_1}, F_{i_2}, k, яка реалізується відповідним оператором (H):

$$H = \begin{cases} Left(I_{i_1}, k-1) \& F_{i_2}, & k \neq 0 \\ Left(I_{i_1}, 0) \& F_{i_2}, & k = 0 \end{cases},\qquad (4)$$

де $Left(Word, Length)$ – функція отримання перших $Length$ символів із слова $Word$; & – конкатенація.

Деяка складність виникає для оберненої задачі (H^{-1}): для існуючого слова знайти три параметри, які його однозначно визначають. Реалізація оператора (4) для бази даних у вигляді недетерміністської функції призводить до того, що неможливо ефективно проіндексувати. Пошук по неіндексованому полю – досить затратний за часом. Для вирішення цієї проблеми запропонований алгоритм (5), суть якого полягає у тому, що вхідне слово розбивається на можливі комбінації незмінної та змінної складових, і пошук проводиться по проіндексованих полях таблиць-множин інфінітивів (I) та флексій (F):

$H^{-1} =$ (5)

```
i=0
l=Довжина(Word)
Повторювати поки i<l
      Head=LEFT(Word,i)
      Tail=RIGHT(Word,l-i)
      Якщо існує інфінітив I_{i1} ∈ I, що починається на
Head та
            існує флексія F_{i2} ∈ F, що дорівнює Tail, то
            визначаємо (I_{i1}, F_{i2}, k)
      Інакше
            i=i+1
Кінець циклу
```

Для жестової мови, у зв'язку з відсутністю у ній словозмін, структура словника буде дещо простішою. У невербальній комунікації мімічна складова відіграє дуже важливу, інколи ключову роль. Варто зазначити, що для синтаксису жестового мовлення для людей з вадами слуху характерна наявність немануального маркера: у

запитальних реченнях, при загальному запитанні – підняті брови, при окремому (приватному) запитанні – опущені брови та нахилена вперед голова. При запереченні – негативні рухи головою, відповідний вираз обличчя. Для розповідних речень характерна наявність на обличчі емоційного забарвлення, що відповідає сенсу інформації, яка передається. Множина жестів подається як (6):

$$Ges = \{Ges_i : Ges_i = \{word\,des, pges \in PGes, em \in Em\}\}, \quad (6)$$

де $word\,des$ – позначення жесту, $PGes \in \{p_1, \ldots, p_{11}, p_{13}, p_{22}\}$, $Em \in \{em_1, em_2, em_3\}$ – множина емоційних забарвлень жестів em_1 – розповідне емоційне забарвлення, em_2 – запитальне емоційне забарвлення, em_3 – інші емоційні забарвлення.

Інфологічну модель української та жестової мови можна відобразити наступним чином (див. рис.1)



Рис.1. Інфологічна модель словників української та жестової мов

У таблиці «Словник української мови» міститься множина слів української мови (W), множина флексій (F) та інфінітивів (I) – у таблицях «Флексії» та «Інфінітиви» відповідно. Для кожного зі слів зберігаються посилання на граматичні категорії, що визначають словозміну: «Множини», «Рід», «УКС» (узагальнена категорія словозміни). Множина жестів міститься в таблиці «Словник жестів» і містить посилання на слово, яким позначається даний жест і тематика, у якій він найчастіше використовується. Множини інфінітивів та жестів містять посилання на «Частини мови» для визначення приналежності їх до певних частин мови.

Для роботи з даною інфологічною моделлю потрібно запропонувати наступну алгоритмічну функціональність (див. (7)-(10)): додавання інфінітива, флексії та слова, отримання інфінітива слова, його флексії, визначення параметрів словозміни слова.

Додавання інфінітива проходить наступним чином:

- *Вхід*: слово-інфінітив та частина мови
- Визначаємо [№ частини мови] з таблиці «Частини мови»
- Додаємо [інфінітив] і [№ частини мови] в таблицю «Частини мови»

(7)

Додавання слова та флексії:

- *Вхід*: слово української мови, інфінітив, множина, рід, узагальнена категорія словозміни
- Визначення [№ множини], [№ роду], [№ узагальненої категорії словозміни], [№ інфінітива] вибіркою з таблиць «Множини», «Рід»,

(8)

```
  «УКС», «Інфінітиви»
• Визначаємо k = позиція слова, починаючи з якої воно відрізняється
  від інфінітива
• l = довжина слова
• флексія = RIGHT(слово,l-i)
• Якщо флексія не порожня
•    Якщо існує [флексія] в таблиці «Флексії» то визначаємо [№
  флексії]
•    Інакше додаємо флексію в таблицю «Флексії» та отримуємо її
  номер
• Кінець, якщо
• Інакше визначаємо номер порожньої флексії з таблиці «Флексії»
• Кінець, якщо
```

Додаємо [№ інфінітива], [№ флексії], [позицію флексії], [№ множини], [№ роду], [№ узагальненої категорії словозміни] в таблицю «Словник української мови».

Отримання інфінітива, флексії, множини, роду, узагальненої категорії словозміни слова:

```
• Вхід: слово української мови
• Визначаємо [№ інфінітива] для слова з таблиці «Словник
  української мови» та [інфінітив] з таблиці «Інфінітиви» по
  зв'язаному полю [№ інфінітива]
• Визначаємо [№ флексії] для слова з таблиці «Словник української
  мови» та [флексію] з таблиці «Флексії» по зв'язаному полю [№
  флексії]
• Визначаємо [№ множини] для слова з таблиці «Словник української
  мови» та [множину] з таблиці «Множини» по зв'язаному полю [№
  множини]
• Визначаємо [№ роду] для слова з таблиці «Словник української
  мови» та [рід] з таблиці «Рід» по зв'язаному полю [№ роду]
• Визначаємо [№ УКС] для слова з таблиці «Словник української
  мови» та [УКС] з таблиці «УКС» по зв'язаному полю [№ УКС]
```
(9)

Отримання частини мови для слова:

```
• Вхід: слово української мови
• Визначаємо [№ інфінітива] для слова з таблиці «Словник
  української мови»
• Для отриманого [№ інфінітива] з таблиці «Інфінітиви» визначаємо
  [№ частини мови] та [частину мови] з таблиці «Частини мови» по
  зв'язаному полю [№ частини мови]
```
(10)

Інфологічна модель пов'язаних узагальнених граматичних конструкцій для перекладу

Після отримання інфологічного подання словників української та жестової мов потрібно побудувати пов'язані узагальнені граматичні конструкції речень української та жестової мов. Під реченням розумітимемо послідовність слів мови та знаків пунктуації. Під граматичною конструкцією речення розумітимемо послідовність слів мови, що відносяться до частин мови, які передають сенс, тобто у речень відкидатимемо службові слова. Вважатимемо, що сенс передається наступними частинами мови: іменники (p_1), числівники порядкові (p_2), числівники кількісні (p_3), числівники

(p_4), числівники типу «два» (p_5), дієслова недоконаного і доконаного виду (p_6), дієслова доконаного виду (p_7), дієслова недоконаного виду (p_8), дієприкметники (p_9), прикметники (p_{10}), займенники (p_{11}), прислівники (p_{13}), дієприслівники (p_{22}). Розрізнятимемо граматичні конструкції за кількістю слів, що в них входять. Елементами граматичних конструкцій (на відміну від речень) будуть множини слів, що зустрічатимуться у реченнях, які складають навчальну вибірку. Отже,

$$Gr = \{Gr_i = \{word_j \in W \| seq_j \in Seq \| p_j \in P, s_j \in S, g_j \in G, i_j \in I, num\}, \\ GStr_i = \{ges_j \in Ges \| gseq \in Gseq \| \ pges \in PGes, num, gesnum\}\} \quad , (11)$$

де *num* – номер елемента в структурі речення української мови, *gesnum* – номер елемента в структурі речення жестової мови,

$$Seq = \{Seq_i : Seq_i = \{word_j \in W : word_j \in W, n\}\},$$

де *n* – номер слова в послідовності,

$$Gseq = \{Gseq_i : Gseq_i = \{ges_j \in Ges : ges_j \in Ges, n\}\},$$

де *n* – номер слова в послідовності.

Після отримання множини граматичних конструкцій проведемо узагальнення кожної із множин елементів, які входять до неї (12). Узагальнення розумітимемо як подання кожної із отриманих множин у вигляді комбінації конкретних граматичних категорій та частин мови:

$$GGr = \{GGr_i = \{ p_j \in P, s_j \in S, g_j \in G, i_j \in I, num\}, \\ GStr_i = \{ pges \in PGes, num, gesnum\}\} \quad . \quad (12)$$

Для реалізації запропонованого підходу було сформовано множину речень, отриманих з програми-комплексу «Українська жестова мова» [6], який використовується в навчальних закладах для нечуючих з метою оволодіння жестовою мовою. Речення в моделі даних було об’єднано в структури, отримані шляхом узагальнення за формулою (11). Структури речень містять замість слів їх послідовності, кожна з яких може використовуватися при побудові речення. Послідовності можуть містити як окремі слова, так і різноманітні множини. Наприклад, речення «він іде», «вона іде», «час йде» було об’єднано в одну послідовність «{він, вона, час, хто, …} іде». Крім того, послідовності можуть містити категорії словозміни (*p,s,g,i*), перетин яких визначає множину слів.

Для подання граматичних конструкцій (11) запропонована наступна інфологічна модель (рис. 2)

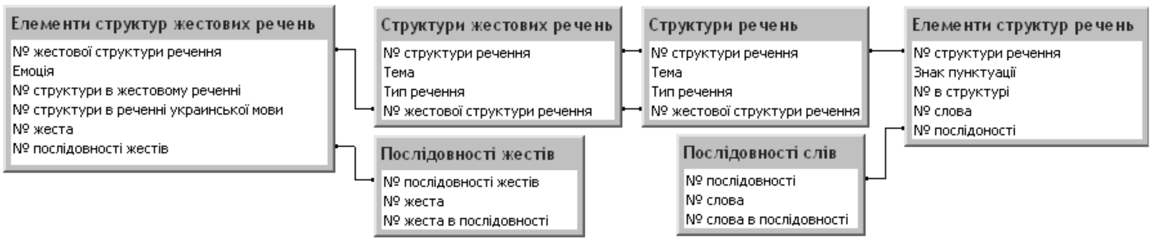


Рис.2. Інфологічна модель зв’язків між структурами речень і структурами жестових речень

Структури речень зберігаються в таблиці «Структури речень», структури жестових речень – в «Структури жестових речень». Дані таблиці містять інформацію про тип речення та тематику речення. «Елементи структур речень» можуть містити як окремі

слова (№ слова), так і послідовності слів (№ послідовності). Для збереження порядку елементів структури речення кожен з елементів має порядковий номер. «Елементи структур жестових речень» також можуть містити жести та послідовності жестів. Для забезпечення відповідності при перекладі також у даній таблиці зберігається порядковий номер елемента структури речення українською мовою.

Для роботи з даною інфологічною моделлю потрібно запропонувати наступну алгоритмічну функціональність (див. (13)-(15)): створення послідовності слів, створення послідовності жестів, створення структури речення, створення структури жестового речення, отримання номера структури речення для речення українською мовою, отримання номера структури жестового речення за номером структури речення українською мовою.

Створення послідовності слів:

```
•  На вхід подається послідовність слів
•  Створюємо запис у таблиці «Послідовності слів», який містить
   перше слово послідовності та його порядковий номер
•  Отримуємо створений [№ послідовності]
•  Для кожного з наступних слів послідовності додаємо
   [№ послідовності], [слово], [№ слова в послідовності] в таблицю
   «Послідовності слів»
```

(13)

Створення послідовностей жестів аналогічне створенню послідовностей слів.

Створення структури речення:

```
•  На вхід алгоритму подається структура речення, яка складається
   зі слів або послідовностей слів, тип та тематика речення
•  Створюємо запис у таблиці «Структура речень», який містить [тип
   речення] та [тему], отримуємо [№ структури речення] для
   створеного запису
•  Для кожного елемента структури речення
•  Якщо елемент – слово, визначаємо [№ слова] з таблиці «Словник
   української мови»
•  Додаємо запис у таблицю «Елементи структур речень», який містить
   [№ структури речення], [№ слова], [Знак пунктуації], [№ в
   структурі]
•  Кінець, якщо
•  Якщо елемент – послідовність:
•     виконуємо пошук її в таблиці «Послідовності слів»
•     Якщо послідовність знайдена, визначаємо [№ послідовності]
•     Інакше створюємо нову послідовність за алгоритмом, описаним
   вище та визначаємо [№ послідовності]
•  Додаємо запис у таблицю «Елементи структур речень», який містить
   [№ структури речення], [№ послідовності], [Знак пунктуації], [№
   в структурі]
•  Кінець, якщо
```

(14)

Створення структури жестових речень аналогічне створенню структури речень, за винятком того, що в таблиці «Елементи структур жестових речень» міститься [№ структури в жестовому реченні] та [№ структури в реченні] для забезпечення відповідності між елементами речень українською та жестовою мовами. Після створення

структури жестового речення, відповідної структурі речення українською мовою, в таблиці «Структури речень» вказується номер відповідної жестової конструкції.

Отримання номера структури для речення українською мовою:

```
• На вхід алгоритму подається речення українською мовою
• Для кожного слова речення знаходимо його № з таблиці «Словник
  української мови»
• Та формуємо індексовану послідовність номерів sentence
• Знаходимо список номерів структур речень list, для яких у
  таблиці «Елементи структур речень» даний sentence[1] міститься в
  послідовності слів або в полі [№ слова] та [№ в структурі] = 1
• Для всіх наступних слів речення серед структур зі списку list
  шукаємо ті, які містять sentence[i] у послідовності слів або в
  полі [№ слова] та  для яких [№ в структурі] = i
• Якщо така структура знайдена – повертаємо її номер
```
(15)

Шляхом аналізу отриманої множини структур речень було отримано множину узагальнених конструкцій української та жестової мов [7], які містять тільки категорії словозміни та не містять окремих слів або множин слів. Для кожного елемента структури речення було виділено множину категорій словозміни (p,s,g,i) та порядок даного елемента в конструкції (11).



Рис.3. Інфологічна модель зв'язків між узагальненими структурами речень та узагальненими жестовими конструкціями

Створення узагальнених конструкцій дозволило на основі аналізу категорій словозміни (p,s,g,i) слів у реченні та його порядкового номера отримувати структуру жестового речення, що йому відповідає.

Інфологічна модель для узагальнених граматичних конструкцій відображена на рис.3.

Таблиця «Узагальнені структури» містить узагальнені конструкції речень української мови. Основними елементами даної таблиці є категорії словозміни p, s, g, i та порядковий номер елемента в структурі речення. Таблиця «Узагальнені жестові конструкції» містить узагальнені конструкції жестових речень, основними елементами яких є частина мови та номер елемента в структурі. Для роботи з даною інфологічною моделлю потрібно запропонувати наступну алгоритмічну

функціональність: отримання номера узагальненої структури для речення української мови, отримання узагальненої жестової конструкції за номером узагальненої структури.

Для створення узагальнених жестових конструкцій запропоновано наступну алгоритмічну функціональність (16):

- Для кожної структури struct з таблиці «Структури речень»
- Для кожної послідовності слів у структурі struct визначаємо параметри словозміни, спільні для її елементів
- Шукаємо узагальнену структуру речення української мови в таблиці «Узагальнені структури», яка містить визначені параметри словозміни
- Якщо такої узагальненої структури не існує, додаємо узагальнену структуру GenStruct у таблицю «Узагальнені структури»
- Кінець, якщо
- Для struct шукаємо відповідну їй структуру gesturestruct з таблиці «Структури жестових речень»
- Визначаємо параметри жестів, спільні для кожного з елементів структури gesturestruct та визначаємо узагальнену жестову конструкцію
- Якщо отримана узагальнена жестова конструкція вже існує, створюємо посилання на неї в GenStruct
- Інакше, у таблицю «Узагальнені жестові конструкції» додаємо узагальнену жестову конструкцію та посилання на неї в GenStruct
- Кінець, якщо

(16)

Крім визначення структури речення необхідною для перекладу є наявність відповідності слів жестам. Використовуючи відповідності між структурами речень та структурами жестових речень, було виділено множину відповідностей «слово → жест». У таблиці «Відповідність слів жестам» міститься номер слова та номер відповідних йому жестів.

Для отримання відповідностей «слово → жест» для кожної структури речень з таблиці «Структури речень» та відповідної їй жестової структури речень з таблиці «Структури жестових речень» було отримано відповідності [7] між послідовностями слів та послідовностями жестів. Використовуючи дані відповідності з таблиць «Послідовності слів» та «Послідовності жестів», було сформовано таблицю «Відповідність слів жестам». У випадку наявності неоднозначних відповідностей «слово → жест» така інфологічна модель не дає можливості отримати чітке відношення між множинами слів та жестів. Якщо одному слову відповідає декілька жестів, для обрання правильного жесту використовуються відношення між реченнями «Структури речень» та «Структури жестових речень» з урахуванням тематики речення, що перекладається. У випадку, якщо кільком словам відповідає один жест, труднощів в отриманні відповідності «слово → жест» не виникає. Крім того, для таких жестів можна отримати список слів-синонімів, які їм відповідають. Заповнення таблиці «Відповідність слів жестам» відбувалося за алгоритмом (17):

- Для кожної структури речення з «Узагальнені структури» визначаємо відповідну їй структуру зі «Структури жестових речень»
- Кожній послідовності слів у структурі речення ставимо у відповідність послідовність жестів

(17)

- Для кожного слова послідовності визначаємо його номер та номер жесту з відповідної жестової послідовності
- Номер слова та номери відповідних жестів додаємо в таблицю «Відповідність слів жестам»

Висновки

У роботі запропоновано інформаційну технологію перекладу з української на жестову мову. Створено інфологічне подання словників української мови як множини інфінітивів, флексій та параметрів словозміни слів, та жестової мови як множини жестів та параметрів, що їх характеризують; розроблено механізм подання відношень між словами української та жестами жестової мов у вигляді узагальнених граматичних конструкцій та зв'язків між жестами і словами; розроблено інформаційну технологію для здійснення автоматизованого перекладу за допомогою отриманих моделей; алгоритмічно та програмно реалізовано розроблену інформаційну технологію для підтвердження її життєздатності. Подальші дослідження спрямовані на створення на основі запропонованої інформаційної технології веб-застосування, яке б дозволило виконувати переклад у режимі online; тестування запропонованої технології більшою кількістю речень; збір статистики використання конструкцій для перекладу.

Література

1. Barmak O. Text to gestures translation for inflected languages / O. Barmak, Iu. Krak, S. Romanyshyn. – LAP LAMBERT Academic Publishing, 2017. – 100 p.
2. Krak Iu. Information technology for automated translation from inflected languages to to sign language /Iu. Krak., V. Kasianiuk., I. Volchyna // Abstracts of XXX International Conference Problems of Decision making Under Uncertainties (PDMU-2017), August 14-19, 2017, Vilnius, Lithuania. – Київ: МП Леся. – 2017. – С. 76-77.
3. Зайцева Г.Л. Жестовая речь. Дактилология: Учеб. для студ. высш. учеб. заведений. – М.: Гуманит. изд. центр ВЛАДОС, 2000. – 61 с.
4. Шульжук К.Ф. Синтаксис української мови / К.Ф. Шульжук. – К. : Академія, 2004. – С. 152–160.
5. Широков В.А. Феноменологія лексикографічних систем / В.А. Широков. – К.: Наук. Думка, 2004. – 327 с.
6. Програма-комплекс "УКРАЇНСЬКА ЖЕСТОВА МОВА" [Електронний ресурс]. – Режим доступу: http://www.mon.gov.ua/education/average/programs_gluh.
7. Крак Ю.В. Побудова множин відповідності для автоматизованого перекладу речень української мови у їх жестові аналоги / Ю.В. Крак, В.С. Касьянюк, І.О.Стеля, І.Л.Черненко // Вісник Київського університету. Серія : фіз..-мат .науки. - 2016. – Вип.1. – С.158 – 161.

Literatura

1. Barmak O. Text to gestures translation for inflected languages / O. Barmak, Iu. Krak, S. Romanyshyn. – LAP LAMBERT Academic Publishing, 2017. – 100 p.
2. Krak Iu. Information technology for automated translation from inflected languages to to sign language /Iu. Krak., V. Kasianiuk., I. Volchyna // Abstracts of XXX International Conference Problems of Decision making Under Uncertainties (PDMU-2017), August 14-19, 2017, Vilnius, Lithuania. - Kyiv: MP Lesya. – 2017. – S. 76-77.
3. Zaitseva G.L. Zestovaya rech. Daktilologia: Ucheb. dlya stud. vyssh. ucheb. zavedenii. – M. Gumanit. izd. centr VLADOS, 2000. – 61 s.
4. Shulzuk K.F. Syntaksys ukrainskoi movy / K.F. Shulzuk. – K.: Akademia, 2004. – S. 152-160.
5. Shyrokov V.A. Fenomenologiya leksykografichnyh system / V.A. Shyrokov. – K.: Nauk. Dumka, 2004. – 327 s.
6. Programa-kompleks "UKRAINSKA ZESTOVA MOVA" [Elektronnyi resurs]. - Rezhym dostupu: http://www.mon.gov.ua/education/average/programs_gluh.
7. Krak Iu. V. Pobudova mnozyn vidpovidnosyi dlya avtomatyzovanogo perekladu rechen ukrainskoi movy u ih zestovi analogy / Iu.V. Krak, V.S. Kasianiuk, I.O. Stelya, I.L. Chernenko // Visnyk Kyiv. universytetu. Ser.: phiz.-mat. nauky. – 2016. – Vyp. 1 .– S. 158-161.

RESUME

Iu.V. Krak, O.V. Barmak, V.S. Kasianiuk, I.O. Stelia

Models of grammar constructions representation for automated translation texts in ukrainian into gestures language

Information technology of translating from Ukrainian into a gesture language is proposed. The basic system of simple sentences structures that have analogues in gesture language and pass semantic constituent information is built. For the system determination mechanism of presentation of relations between words in Ukrainian and gestures in gesture language has been worked out. It allowed to put corresponding gestures in accordance to the words, mechanism of presentation of the generalized grammatical constructions of translation from Ukrainian into gesture language that allowed by determination of grammatical structure of an introductory sentence to find the corresponding structure of gesture sentence for translation. Using parameters of incoming sentence words replacement to determine its grammatical structure is the approach advantage.

Ukrainian grammatical dictionary (belongs to flexional languages) on the basis of formal model of word-replacement has been built, i.e. linguistic criteria according to that all set of words of language is broken up on certain subsets with empty mutual crossing, words-replacement into each of them takes place on the same rules, has been set and formalized. The subsets of words with such properties are called as word-replacement paradigm types. They contain lexemes group, with word-replacing paradigm characterized by the identical amount of grammatical forms, word-replacement into the group takes place by the same (the only one) rule. Seven paradigm types have been distinguished and investigated for Ukrainian. Base word-form number out of corresponding set of base word-forms instead of complete text of words is kept in database, number of position in a word, up to which the base word-part is unchanging, and number of flexion (word part variable) from a corresponding set. Due to the absence of word changes in gesture language dictionary structure consists of set of gestures and no-manual markers that pass mimic and emotional gesture constituent.

The translation method is built by receipt mutually bijective constructions pairs which pass sense: sentence in input language into a corresponding sentence in gesture language. Structures (generalizations) containing words set (word-replacement parameters given) used on the fixed places instead of certain words in sentence have been gotten by fixing incoming sentence word order. The Sequences can contain separate words as well as word sets. The sequences can be changed and complemented in translation process. Analogical structures are built for gesture sentences as well.

The approach has been approved on considerable number of sentences taken from Ukrainian in gestures program for the specialized schools. About one hundred summarizing constructions are gotten as a result.

Further researches follow the system is to be additionally extended by greater amount of sentences (from everyday life for different age-related and professional groups) that will allow generalization of great number of typical sentences sufficient for everyday communication to get as well as arbitrary text information automated translating into language of fingers to be built up.

Надійшла до редакції 21.02.2018

